DOCUMENT RESUME

ED 170 382 TM 009 243

AUTHOR Berk, Ronald A.
TITLE A Critical Review of Content Domain Specification/Item Generation Strategies for Criterion-Referenced Tests.
PUB DATE Apr 79
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979) ; Table 1 marginally legible due to small type

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Behavioral Objectives; Comparative Analysis; *Criterion Referenced Tests; Evaluation Criteria; *Specifications; *Test Construction; *Test Items; Test Reliability

ABSTRACT

As alternatives to the objectives-based approach to specifying content domains for test construction purposes, six strategies are proposed: (1) amplified objectives; (2) Instructional Objectives Exchange (IOX) test specifications; (3) item transformations; (4) item forms; (5) algorithms; and (6) mapping sentences. Their effectiveness is assessed in terms of the extent to which they provide an unambiguous domain definition and explicit rules for constructing items such that any two test-makers would produce identical items from the same specifications. A rating system was also devised to evaluate the strategies according to eight factors: clarity, simplicity, availability, development time, development cost, adaptability, domain appropriateness, and practicability. Item transformations, item forms, and algorithms are judged to offer the most rigorous and precise specifications, while amplified objectives, IOX test specifications, and mapping sentences tend to be the most practical. (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED170382

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A CRITICAL REVIEW OF CONTENT DOMAIN SPECIFICATION/ ITEM GENERATION STRATEGIES FOR CRITERION-REFERENCED TESTS

Ronald A. Berk

The Johns Hopkins University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ronald A. Berk

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the annual meeting of the American Educational Research Association

San Francisco, April 1979

TM009 243

Abstract

Six strategies for specifying content domains have been proposed as alternatives to the objectives-based approach of the 1960's and early 1970's. Their effectiveness is assessed in terms of the extent to which they provide an unambiguous domain definition and explicit rules for constructing items such that any two test-makers would produce identical items from the same specifications. A rating system was also devised to evaluate the strategies according to eight factors of practicability. Item transformations, item forms, and algorithms offered the most rigorous and precise specifications, while amplified objectives, IOX test specifications, and mapping sentences tended to be the most practical.

A Critical Review of Content Domain Specification/Item Generation Strategies for Criterion-Referenced Tests

In recent years, some of the leading proponents of criterion-referenced tests have argued that the objectives-based approach to specifying content domains for purposes of test construction is inadequate. They contend that a content outline, set of instructional and behavioral objectives, and table of specifications or similar "blueprint" tend to result in an ambiguous definition of a domain (Brennan, 1974; Hambleton, Swaminathan, Algina, & Coulson, 1978; Millman, 1974; Nitko, 1974; Popham, 1978). The arguments focus on the subjectivity involved in composing those specifications. That is, the selection of content topics and objectives is quite arbitrary, typically representing only one test-maker's conceptualization of the domain. Potentially these specifications are open to different interpretations by different test-makers. Given this ambiguity, the specifications are felt to be inadequate for item writing.

Along with this criticism it has been charged that traditional item construction procedures used to write items from the specifications are also ambiguous. The critics argue that employing these procedures results in a set of items that reflects the biases and idiosyncrasies of each test-maker. Consequently, different test-makers would likely develop different items from the same specifications.

In order to overcome these deficiencies, several other strategies have been recommended. They attempt to provide an unambiguous definition of a domain and explicit rules for generating items such that any two test-makers would produce identical items from the same specifications. The extent to which this goal is achieved, as well as the method, provides the focus of this paper.

Six alternative strategies for domain specification were critically reviewed: (1) amplified objectives; (2) IOX test specifications; (3) item transformations; (4) item forms; (5) algorithms; (6) mapping sentences.

Method

First, the general and technical characteristics of the strategies were surveyed. There was a particular emphasis on the major components, the rule structure by which the content domain is linked to the test items, the type of item domain from which the item sample is generated, and the projects and content domains to which the strategy has been applied. A comparison based on these characteristics followed.

Second, a rating system with criteria was devised to present a profile of the advantages and disadvantages of each strategy. This evaluation concentrated on eight factors related to practicability: (1) clarity; (2) simplicity; (3) availability; (4) development time; (5) development cost; (6) adaptability; (7) appropriateness for the content domains of reading, language, mathematics, science, social studies, affective, and psychomotor behaviors; (8) practicability for teachers, district evaluators, and state evaluators. These factors are defined below:

1.	Clarity	The extent to which two item writers would produce similar items from one set of specifications.
2.	Simplicity	The amount and level of training required to effectively develop and use the specifications.
3.	Availability	The extent to which the specifications are readily available from commercial or other sources.
4.	Development time	The amount of time needed to complete the specifications as a function of a practitioner's estimated available time.
5.	Development cost	The cost related to 2 and 4.

6.	Adaptability	The extent to which the strategy can be modified for use in instructional programs of varying lengths (e.g., three weeks, ten months), grade levels (K - college), and content orientations (e.g., reading, functional skills for learning disabled children).
7.	Domain appropriateness	The extent to which the usefulness of the strategy for defining specific content and behavioral domains has been effectively demonstrated and tested. Isolated small-scale applications are indicated with a "-".
8.	Practicability	The extent to which the strategy is capable of being implemented by teachers, district evaluators, and state evaluators given the necessary resources. This factor is a function of factors 2, 4, 5, and 6.

Results and Conclusions

The examination of the characteristics of the six strategies revealed that the extent to which they achieve an unambiguous domain definition and overcome the stated deficiencies of objectives varies considerably. Amplified objectives, IOX test specifications, and mapping sentences fall short of these goals. This is due in large part to their reliance on traditional item construction rules. They might be characterized as achieving a relatively less ambiguous definition. Item transformations, item forms, and algorithms attain unambiguous definition by means of sophisticated rule structures. When short-cut approaches that modify the structures

were employed (e.g., Lowerre & Scandura, 1973; Roid & Haladyna, 1978a, 1978b), the "unambiguity" was compromised.

Insert Table 1 about here

The profiles of the strategies suggested that the rigor and precision of the specifications are inversely related to their practicability. Item transformations, item forms, and algorithms which are capable of generating finite item domains and domain sampling seemed to be the most impractical. They should be considered as computer-based technologies (cf. Millman, in press). Amplified objectives, IOX test specifications, and mapping sentences which are associated with the conceptualization of an infinite item domain indicated the greatest potential for use by teachers and evaluators. Unfortunately, all of the strategies except amplified objectives and IOX test specifications have been effectively demonstrated in only one content domain. It is either reading (item transformations), mathematics (item forms, algorithms), or affective behaviors (mapping sentences).

Insert Table 2 about here

Certainly a great deal of research needs to be conducted to appraise the validity of these "alternative" strategies for specifying domains in the other content areas. The findings of this review should furnish specific directions for that research.

References

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Baker, E. L. Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. Educational Technology, 1974, 14, 10-16.

Berk, R. A. The application of structural facet theory to achievement test construction. Educational Research Quarterly, 1978, 3, 62-72.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Brennan, R. L. The evaluation of mastery test items (Final Report, Project No. 2B118). Washington, D.C.: National Center for Educational Research and Development, U. S. Office of Education, 1974.

Castro, J. G., & Jordan, J. E. Facet theory attitude research. Educational Researcher, 1977, 11, 7-11.

Finn, P. J. A question writing algorithm. Journal of Reading Behavior, 1975, 7, 341-367.

Guttman, L. Integration of test design and analysis. In Proceedings of the 1960 invitational conference on testing problems. Princeton: Educational Testing Service, 1970.

Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. B. Criterion-referenced testing and measurement. A review of technical issues and developments. Review of Educational Research, 1978, 48, 1-47.

Hively, W., Maxwell, G., Rabehl, G., Sension, E., & Lundin, S. Domain-referenced curriculum evaluation: A technical handbook and a case study from the MINNEMAST Project (CSE Monograph Series in Evaluation, No. 1). Los Angeles: Center for the Study of Evaluation, University of California, 1973.

Lowerre, G., & Scandura, J. M. Conceptually based development and evaluation of individualized materials for critical reading based on logical inference. Reading Research Quarterly, 1973, 9, 186-205.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current practices. Berkeley, Calif.: McCutchan, 1974. Chap. 6.

Millman, J. Computer-based item generation. In R. A. Berk (Ed.), Criterion-referenced measurement: The state of the art. Baltimore, MD: The Johns Hopkins University Press, in press.

Nitko, A. J. Problems in the development of criterion-referenced tests: The IPI Pittsburgh experience. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in Evaluation, No. 3). Los Angeles: Center for the Study of Evaluation, University of California, 1974. Pp. 59-82.

Osburn, H. G. Item sampling for achievement testing. Educational and Psychological Measurement, 1968, 28, 95-104.

Popham, W. J. Selecting objectives and generating test items for objectives-based tests. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in Evaluation, No. 3). Los Angeles: Center for the Study of Evaluation, University of California, 1974. Pp. 13-25.

Popham, W. J. Criterion-referenced measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Popham, W. J. Domain specification strategies. In R. A. Berk (Ed.), Criterion-referenced measurement: The state of the art. Baltimore, MD: The Johns Hopkins University Press, in press.

Roid, G. H., & Haladyna, T. M. A comparison of several linguistic-based, multiple-choice item writing algorithms. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March, 1978.

Roid, G. H., & Haladyna, T. M. Comparison of objectives-based and modified-Bormuth item writing techniques. Educational and Psychological Measurement, 1978, 38, 19-28.

Scandura, J. M. Structural learning I: Theory and research. New York: Gordon and Breach, 1973.

Scandura, J. M. Problem solving: A structural process approach with instructional implications. New York: Academic Press, 1977.

Table 1

General and Technical Characteristics of Six Strategies for Content Domain Specification/Item Generation

Strategy	Developer(s)	Project Application/Domain(s)	Rule Structure	Major Components	Item Domain
Amplified Objectives	Popham (1974) Baker (1974)	IOX Test Development/Reading, Language Mathematics, Social Studies	Traditional Item Construction Rules	Objective Sample Item Testing Situation Response Alternatives Criterion for Correctness	Infinite
IOX Test Specifications	Popham (1978, in press)	IOX Test Development/Reading, Language Mathematics, Social Studies, Affective	Traditional Item Construction Rules	General Description Sample Item Stimulus Attributes Response Attributes Specification Supplement	Infinite
Item Transformations	Bormuth (1970) Anderson (1972) Finn (1975)	UCLA Readability Project/Reading	Transformational Rules	Base and Derived Sentence	Finite
Item Forms	Hively et al. (1973) Osburn (1968)	MINNEMAST Project/Mathematics	Generation Rules	Shell Replacement Sets	Finite
Algorithms	Scandura (1973, 1977)	MERG Projects[a]/Mathematics, Reading	Rules of Competence (Higher and lower order)	Equivalence Classes	Finite
Mapping Sentences	Guttman (1970) Castro & Jordan (1977) Berk (1978)	Cross Cultural Research/Attitudes	Facet Design and Item Construction Rules	Facets Facet Elements Semantic Profiles	Infinite

[a] The Mathematics Education Research Group (MERG) Projects provided the bases for most of the empirical research on algorithms.

Table 2

Ratings of Seven Strategies for Content Domain Specification in Terms of Eight Practicability Factors[a]

Factor	Objectives	Amplified Objectives	IOX Test Specifications	Item Transformations	Item Forms	Algorithms	Mapping Sentences
1. Clarity	-	+	+	+ +	+ +	+ +	+
2. Simplicity	+ +	+	+	- -	- -	- -	+
3. Availability	+ +	+	+	- -	- -	- -	- -
4. Development time	+	- -	- -	- -	- -	- -	+
5. Development cost	+	- -	- -	- -	- -	- -	+
6. Adaptability	+ +	+ +	+ +	- -	- -	-	+ +
7. Domain Appropriateness							
a. Reading	+ +	+ +	+ +	+ +	- -	-	-
b. Language	+ +	+ +	+ +	- -	- -	- -	- -
c. Mathematics	+ +	+ +	+ +	- -	+ +	+ +	-
d. Science	+ +	- -	- -	- -	+	- -	-
e. Social studies	+ +	+ +	+ +	- -	- -	- -	- -
f. Affective	+ +	-	+	- -	-	- -	+ +
g. Psychomotor	+ +	- -	- -	- -	- -	- -	- -
8. Practicability							
a. Teachers	+	- -	- -	- -	- -	- -	+
b. District evaluators	+ +	-	-	- -	- -	- -	+ +
c. State evaluators	+ +	-	-	- -	- -	- -	+ +

[a] Rating System: + + Marked advantage
+ Slight advantage
- Slight disadvantage
- - Marked disadvantage